DOCUMENT RESUME

ED 058 312 TM 001 027

AUTHOR	Harris, Chester W.
TITLE	An Interpretation of Livingston's Reliability Coefficient for Criterion-Referenced Tests.
INSTITUTION	California Univ., Santa Barbara.
PUB DATE	17 Mar 71
NOTE	7p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Analysis of Variance; *Criterion Referenced Tests; *Reliability; Standard Error of Measurement; *Statistical Analysis; *Test Reliability; *Tests; Tests of Significance; True Scores
IDENTIFIERS	*Livingston's Reliability Coefficient

ABSTRACT

Livingston's work is a careful analysis of what occurs when one pools two populations with different means, but similar variances and reliability coefficients. However, his work fails to advance reliability theory for the special case of criterion-referenced testing. See ED 042 802 for Livingston's paper. (MS)

AN INTERPRETATION OF LIVINGSTON'S RELIABILITY COEFFICIENT FOR CRITERION-REFERENCED TESTS

Chester W. Harris

University of California

Santa Barbara

March 17, 1971

It may be of value to present some comments about a recent development in reliability theory [Livingston, 1971] even though Livingston's material is, as of this writing, in unpublished form. The reason for presenting these comments now is that his work has become known to a number of people and is being used for reliability estimation of so-called criterion referenced tests. For example, Hsu [1971] in a paper presented to the 1971 AERA Convention, reports Livingston reliabilities for tests used at the University of Pittsburgh. These comments are intended to clarify the meaning of a Livingston reliability coefficient.

I shall not comment in detail on the Livingston paper. His algebra is impeccable, and the formulas he derives follow exactly from his definitions and from the assumptions of classical reliability theory. Instead, I shall show that a Livingston reliability coefficient, given as

$$K^2(X,T_x) = \frac{\sigma^2(T_x) + (\mu_x - C_x)^2}{\sigma^2(X) + (\mu_x - C_x)^2} \quad ,$$

is identical with a conventional reliability coefficient when that coefficient is based on <u>two</u> populations with means equally distant above and below C_x. For this to be true it is necessary that $\sigma^2(T_x)$ and $\sigma^2(E)$ be identical in the two populations, which thus implies that the conventional reliability coefficient for either population:

$$\frac{\sigma^2(T_x)}{\sigma^2(T_x) + \sigma^2(E)}$$

is identical in the two populations. Also, the classical assumption of independence of true score and error must hold in both populations. μ_x may be taken as the mean of either population, and then C_x, which Livingston designates as a criterion score, must be the mean of the two populations pooled.

Given two populations with $\sigma^2(X_1) = \sigma^2(X_2) = \sigma^2(X)$, with $\sigma^2(T_{x_1}) = \sigma^2(T_{x_2}) = \sigma^2(T_x)$, and with $\sigma^2(E_1) = \sigma^2(E_2) = \sigma^2(E)$, but with means $\mu_1 \neq \mu_2$, we can write the variance of observed scores for the pooled populations as

$$\sigma^2(X) + \frac{\mu_1^2}{2} + \frac{\mu_2^2}{2} - \frac{(\mu_1 + \mu_2)^2}{4}$$

This follows by determining the expected value of the squared deviations of the scores in the two populations from μ_p, which is the mean of the pooled (equally weighted) populations and is identical with $\frac{\mu_1 + \mu_2}{2}$.

Now $\frac{\mu_1^2}{2} + \frac{\mu_2^2}{2} - \frac{(\mu_1 + \mu_2)^2}{4}$ has an equivalent form equal to either $(\mu_1 - \mu_p)^2$ or $(\mu_2 - \mu_p)^2$, both of which have the same value when μ_p lies half way between μ_1 and μ_2. Consequently the variance of the two populations pooled is

$$\sigma^2(X) + (\mu_1 - \mu_p)^2 .$$

Since classic true scores have the same mean as observed scores, by a similar argument the variance of true scores in the two populations pooled is $\sigma^2(T_x) + (\mu_1 - \mu_p)^2$. Therefore the classic reliability coefficient based on these <u>two</u> populations pooled is

$$\frac{\sigma^2(T_x) + (\mu_1 - \mu_p)^2}{\sigma^2(X) + (\mu_1 - \mu_p)^2}$$

which is Livingston's coefficient when we identify his μ_x with our μ_1 and his C_x with our μ_p.

Livingston offers this illustration or interpretation of his coefficient. Suppose an employer wishes to employ workers who score at least 75 on a given test. The criterion score, C_x, is set in advance as 75. If the employer then tests a population with mean score, μ_x, of 60 he is likely to reject many of these persons. Livingston would, I believe, argue then that his reliability coefficient which incorporates the quantity $(60 - 75)^2$ in both the numerator and the denominator of the formula is the correct indicator of the confidence we can have in believing that an individual in this population has a true score less than 75.

Let us make two points.

If we have available a population with a mean of 60 and another with a mean of 90, both populations identical in the variance of observed, true, and error scores, the conventional reliability coefficient based on the two populations pooled will be

$$\frac{\sigma^2(T_x) + (60 - 75)^2}{\sigma^2(X) + (60 - 75)^2}$$

which is the Livingston coefficient based on the population with mean of 60 when C_x is set at 75. Apparently then the Livingston coefficient requires that in addition to the population with mean of 60 one must regard as reasonable the postulation of a similar population with mean 15 points <u>above</u> the C_x of 75, or a mean of 90. One can readily conceive situations in which this is not reasonable. Suppose the test has a ceiling of, say, 20 points and the criterion is set at 16 points. If we use this test with a population having a mean of 10, then there can be no population with a mean of 22, 16 plus (16-10), such that the pooling of the two populations gives a conventional reliability coefficient equal to the Livingston coefficient. Both "floors" and "ceilings" on tests are facts of life and they may be inconsistent with the procedure Livingston recommends. In general, a reliability coefficient depends upon the range of talent. Theoretically at least, one can manipulate the magnitude of the reliability coefficient by altering the range of talent. Livingston appears to have secured his "bigger" coefficients as a consequence of implicitly extending the range of talent.

The second point is that the variance of the errors of measurement may be a better guide to one's confidence in asserting that a given individual has a true score below (or above) a certain point than is the magnitude of the reliability coefficient. Again, at least theoretically, manipulating the magnitude of the reliability coefficient by altering the range of talent does not alter the variance of the errors of measurement. Textbooks commonly state that the standard error of measurement is independent of the range of talent. Now it is almost painfully obvious that

$$\sigma^2_{(X)} + (\mu_x - C_x)^2 \left[1 - \frac{\sigma^2_{(T_x)} + (\mu_x - C_x)^2}{\sigma^2_{(X)} + (\mu_x - C_x)^2} \right]$$

equals $\sigma^2_{(X)} - \sigma^2_{(T_x)}$, which is $\sigma^2_{(E)}$ or the variance of the errors of measurement. Thus although Livingston's reliability coefficient is larger than the conventional one, the standard error of measurement is the same, and consequently this larger coefficient does <u>not</u> imply a more dependable determination of whether or not a true score falls below (or exceeds) a given criterion value.

Viewed from this standpoint, Livingston's work appears to be primarily a careful spelling out of what occurs when one pools two populations with different means, but similar variances and (conventional) reliability coefficients. If this view is correct, we must conclude that his work fails to advance reliability theory for the special case of criterion referenced testing.

REFERENCES

Hsu, Tse-Chi. Empirical Data on Criterion-Referenced Tests. Paper presented at the annual meeting of the American Educational Research Association, New York City, 1971.

Livingston, Samuel A. Criterion-referenced Applications of Classical Test Theory. Unpublished paper, Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Maryland. n.d.

dd